AD_____

GRANT NUMBER DAMD17-94-J-4328

TITLE: Computer-Aided Diagnosis and Automated Screening of
Digital Mammogram

PRINCIPAL INVESTIGATOR: Kevin S. Woods, Ph.D.

CONTRACTING ORGANIZATION: University of South Florida
                          Tampa, Florida 33620-7900

REPORT DATE: October 1997

19980526 069

TYPE OF REPORT: Final

PREPARED FOR: Commander
               U.S. Army Medical Research and Materiel Command
               Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
                        distribution unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

DTIC QUALITY INSPECTED 2

i

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE October 1997	3. REPORT TYPE AND DATES COVERED Final (1 Sep 94 - 29 Sep 97)

4. TITLE AND SUBTITLE
Computer-Aided Diagnosis and Automated Screening of
Digital Mammogram

5. FUNDING NUMBERS
DAMD17-94-J-4328

6. AUTHOR(S)

Kevin S. Woods, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of South Florida
Tampa, Florida 33620-7900

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

We have developed image analysis software which is capable of
detecting mammographic abnormalities which would be used in a
second reader scenario to prompt a radiologist to more carefully
analyze suspicious regions in the mammogram. For an average of
about 2 prompts per image, our algorithm detected 70% of the
lesions in our database. In a second reader scenario, it is only
necessary to detect a lesion in at least one of the two views.
Using this criteria, our algorithm detected 90% of the masses,
and perhaps more importantly 97% of the malignant masses were
found. The automated prompting and the additional information
provided by computerized image analysis should result in greater
repeatability and uniformity in the standard of care. As several
recent studies have indicated, it should also result in some
increase in sensitivity for a given level of specificity. The
extra cancers detected would then be treated earlier and less

14. SUBJECT TERMS
Breast Cancer, computer-aided diagnosis, digital
mammography

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

_____ 10/27/97
PI - Signature   Date

Table of Contents

COMPUTER-AIDED DIAGNOSIS AND AUTOMATED SCREENING OF DIGITAL MAMMOGRAMS

Kevin Woods

Dept. of Computer Science & Engineering
University of South Florida, Tampa, FL 33620-5399

FINAL REPORT - 10/29/97

1 Introduction

Breast cancer is by far the most common cancer among women. Although lung cancer has a lower incidence (fewer diagnosed cases) than breast cancer, more women die each year of lung cancer. However, the majority of deaths from lung cancer can be attributed to smoking, and so breast cancer continues to be the leading cause of *nonpreventable* cancer death. Current statistics indicate that 1 in 9 women will develop breast cancer at some time in their life [1]. Additionally, the etiologies of malignant breast cancer are unclear, and no single dominant cause has emerged. Although there is currently no known way of preventing breast cancer, the earlier a cancer is detected and treated, the better the prognosis [2]. Currently, X-ray mammography is the single most important factor in early detection, and screening mammography could result in at least a 30 percent reduction in breast cancer deaths [2].

Due to a variety of factors, accurate interpretation of mammograms is considered quite difficult. Studies have shown that screening suffers from large variability in detection rates [3, 2]. There is a psychovisual phenomenon that applies to mammographic interpretation that guarantees a radiologist will occasionally fail to perceive significant abnormalities [4]. This is supported by studies showing that radiologists do not identify all breast cancers that are visible on retrospective review, and that many malignant abnormalities are not recommended for biopsy [5, 6, 7, 8, 9]. In one study, as many as 10% of true malignancies were missed because they were overlooked [7]. Another study demonstrated that approximately 30 percent of lesions will be visible in a mammogram but missed for some reason, and another 30 percent of lesions will have subtle signs of malignancy that are difficult to detect [2]. Therefore, steps taken towards increasing the reliability and consistency of mammogram interpretation will have a significant and direct positive impact on early detection of breast cancer.

The purpose of this research is to develop computer software for the task of interpreting digital mammogram images. Computer-aided diagnosis (CAD) is recognized to hold great promise for improving the sensitivity, specificity, and cost-effectiveness of mammography. Image analysis techniques can potentially be used to help the radiologist to interpret mammograms with greater accuracy, reliability, repeatability, and efficiency than would otherwise be possible. We envision a "second reader scenario," in which all mammograms are still read by the radiologist in a manner similar to current practice. In addition, computerized image analysis is used to suggest possible suspicious regions in the image so that the radiologist can then examine these regions more carefully. In effect, the computerized image analysis provides benefits similar to an independent reading of the mammogram by a second radiologist.

As several recent studies have indicated the use of CAD in mammography screening is feasible and beneficial [10, 11, 12, 13, 14]. It will result in some increase in sensitivity for a given level of specificity; that is, fewer missed cancers with the same biopsy rate. The cancers detected as a result of the increased sensitivity would then be treated earlier and thus less expensively and with a higher cure rate. As a specific example, the study by Kegelmeyer et al. [10] demonstrated that use of a CAD tool increased the average true positive detection rate of participating radiologists from 80.6% to 90.3% without any increase in their average false positive detection rate.

2 Body

This section describes in detail the approach we have developed for mammogram image analysis. We begin with an overview of the feature extraction, classification, and image processing algorithms. Next, we discuss the the experimental methods, data, and any assumptions, including the train/test protocol and performance evaluation criteria utilized, and the results are presented. Finally, we draw some conclusions and make several recommendations for other researchers.

2.1 Algorithm Overview

The basic algorithm framework involves two phases, a sophisticated pixel-level segmentation, and a simple region-level classification. Segmentation at the pixel level encompasses several elementary steps. First, a set of seven features is computed at every pixel in the breast region of the mammogram image. These features include some general purpose measures of local image texture, and more complex features specifically engineered to respond to characteristics associated with mammographic abnormalities. Next, statistical classification is used to assign each pixel with a probability of suspiciousness. This process is depicted in Figure 1. The result is called a probability image which is smoothed and thresholded to produce a binary template which denotes which pixels are considered suspicious and which are thought to be normal. Finally, suspicious pixels are organized into regions by grouping connected pixels. At the end of this stage of processing, we have blobs denoting the locations of suspicious regions in the mammogram image.

The region-level classification step is simply one of removing very small regions in the

1) Input Image

2) Feature Extraction

Texture Energy | Fractal Texture | Local Contrast | Spicule Finder | Circle Finder | Oval Finder

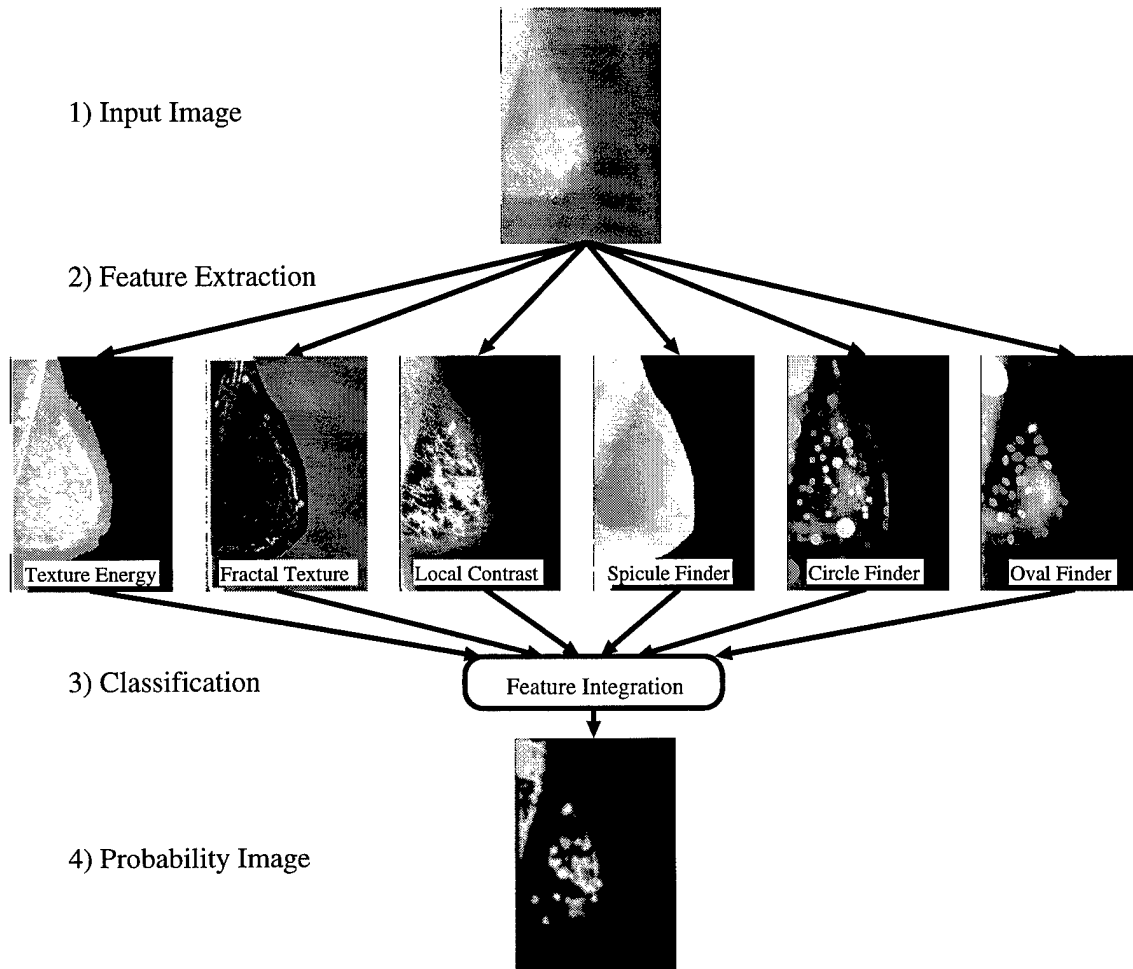3) Classification

Feature Integration

4) Probability Image

Figure 1: An overview of the algorithm used for mammogram image analysis. This example shows a mammogram image and the results of six feature extraction routines displayed as feature images. The feature images are input to a classifier, and the result is a probability image.

binary template which are the result of the previous segmentation stage. Since our goal is to prompt a radiologist to more closely examine any suspicious locations in the mammogram, cross-hairs are overlaid on the original mammogram image at locations which correspond to the centroids of any regions remaining in the binary template. Figure 2 shows the final output of our algorithm along with the ground truth for an example mammogram image.

2.1.1 Pixel-Level Features

In some of our previous work [15], we discussed the benefits of concentrating development efforts and algorithm sophistication on the pixel-level analysis. Indeed, the features extracted at the pixel level are a major factor in overall system performance. Therefore, a large portion of the development effort was directed towards development and refinement of feature extraction routines.

3

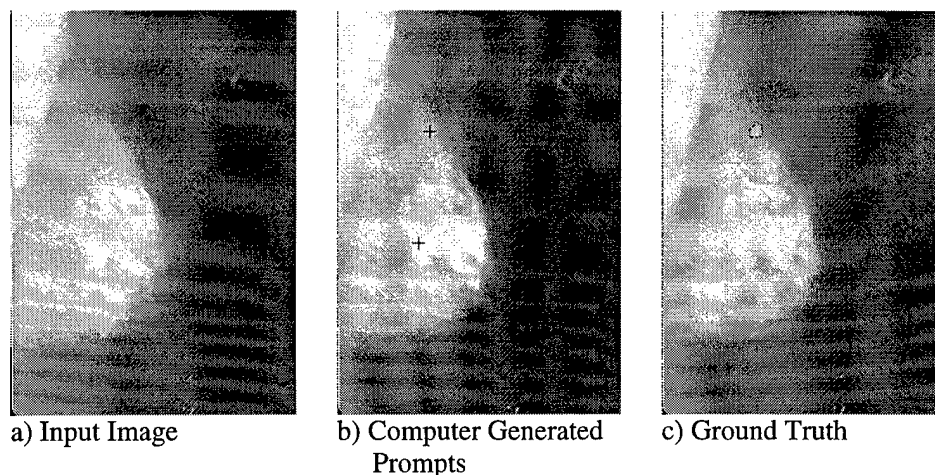a) Input Image   b) Computer Generated   c) Ground Truth
Prompts

Figure 2: (a) An example mammogram image, and (b) the output of our detection algorithm compared with (c) the ground truth as marked by an expert mammopgrapher.

After examining and evaluating over a hundred different types of features and feature variations, we eventually settled on seven features for mass detection, which are described in the following paragraphs.

■ Mass region: template value = 1

▨ "Don't care region": template value = 0

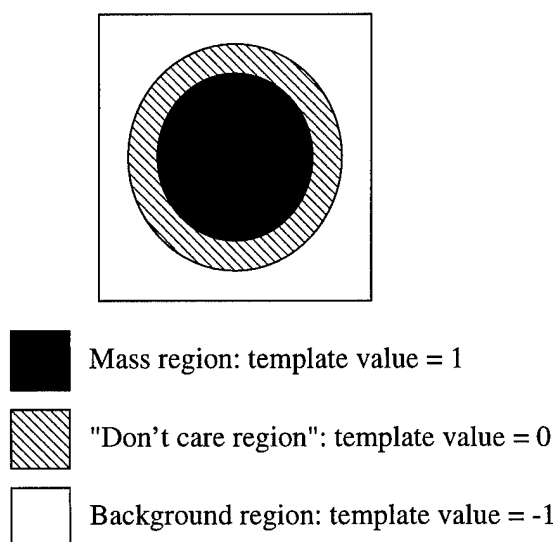□ Background region: template value = -1

Figure 3: The circle template used to detect circular masses.

Multi-scale circle template matching. Masses are more dense than surrounding breast tissue, and many are roughly circular in shape. Therefore, a feature that responds to circular densities is useful. Lai et al [16] describe a normalized cross-correlation measure for matching a circle template to a mammogram image. The 11 pixel by 11 pixel template we use is shown in Figure 3. All pixels within a seven pixel radius of the center pixel set to one (1). A "don't care" region is defined by setting the template to zero (0) for a ring of pixels with Euclidean distances from 8 to 9 from the center. The "don't care" region in the

4

template permits accurate matches for masses that are not perfectly circular. Pixels outside the "don't care" region are set to minus one (-1). The normalized cross-correlation measure returns a value in the range [-1,1], and a bright circle on a dark background would have a value in the range (0,1]. To account for different size masses, the template matching is performed on images scaled to various spatial resolutions. Images are scaled such that the 7 pixel radius of the circle template corresponds to masses which range in size from 6 mm to 46 mm in diameter (in increments of 2 mm). For each pixel, we keep the value of the strongest match (highest normalized cross-correlation measure), and note the metric radius of the template corresponding to the strongest match. A post-processing step assigns all pixels within the best matching circle regions the same value as the pixel located at the centroid of the template. These circle regions are not permitted to overlap. So we start with the pixel with the strongest template matching measure in the image, and 'fill in" the circular region around it with the value of the center pixel. As noted previously, the size of the circular region is known as a result of the multi-scale template matching. We then find the next strongest matching pixel, and repeat the fill-in procedure until no more circles can be found which do not overlap any previously filled circle regions. Any pixels that are not part of a filled-in circle simply retain the value returned from the multi-scale template matching. Example results for the multi-scale circle template matching algorithm are shown in Figure 4.
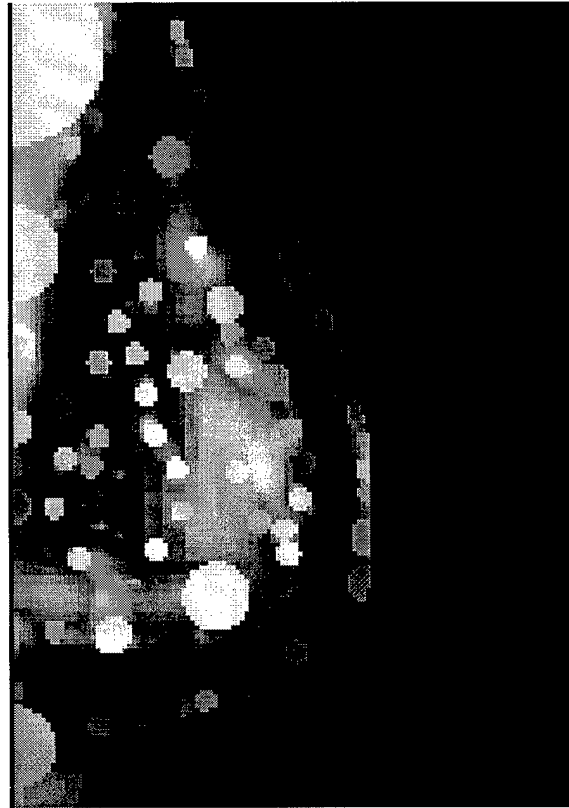

Figure 4: Example results for the multi-scale circle template matching algorithm.

Multi-scale oval template matching. Since many masses are more oval-shaped, we

applied the multi-scale template matching algorithm using an oval template rotated to four different orientations (see Figure 5). The oval templates were created by scaling a circle template with an inner radius of 5 pixels by a factor of 1.5 in four directions (0, 45, 90, -45 degrees). Template matching was performed at resolutions corresponding to ovals with semi-minor axis lengths of 3 mm to 17 mm in increments of 2 mm. Again, the semi-major axis is 1.5 times the semi-minor axis. As before, the results of the template matching procedure are post-processed by assigning all pixels within the best matching oval regions the same value as the pixel located at the centroid of the template.

Mass region: template value = 1

"Don't care region": template value = 0
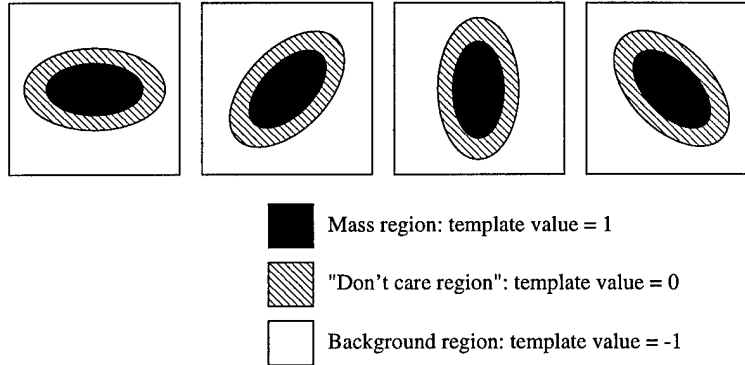
Background region: template value = -1

Figure 5: An oval template rotated to 0, 45, 90, and -45 degrees is used to detect oval-shaped masses.

Multi-scale orientation analysis. The Analysis of Local Edge Orientation (ALOE) measurement was originally used by Kegelmeyer [17, 18, 10] for the detection of spiculated lesions. His original ALOE measurement is computed for each pixel by centering a $4cm$ by $4cm$ window on each pixel, and extracting a histogram of the edge orientations across the window. The histogram is normalized by dividing each bin height by the total number of pixels in the histogram. Finally, the ALOE feature is computed as the standard deviation of the histogram bin heights. The idea here is that a pixel in the area of a spiculated lesion will have a low ALOE value since there will be edges (i.e. the spicules) radiating in all directions causing a relatively flat histogram. The edge orientations for our ALOE measurements are derived from the output of a steerable filter algorithm which utilizes second derivative of Gaussian filters [19]. To detect different size lesions, the ALOE histogram is computed for several window sizes. Only the minimum value over all window sizes is retained.

Relative contrast. A measure of how bright a pixel is compared to surrounding pixels is the relative contrast. This measure is computed for a pixel by subtracting the mean of a 4cm by 4cm window (centered on the pixel) from the pixel value, and dividing the result by the pixel value.

Fractal-based texture measure. A texture feature based on fractal dimension is computed using the techniques described in [20, 21]. Basically, a surface area measurement for each pixel is computed at several different scales, or resolutions. Linear regression is performed on a set of points, where the x value is the logarithm of the scale value, and the y value is the logarithm of the area measurement. At each pixel, a regression derived feature is computed as the slope of the fitted line.

6

Laws L7*E7 texture energy measure. Laws' texture energy measures [22] involve the application of a set of convolution kernels to an image. Each kernel is designed to respond to a different local property (i.e. texture). A texture energy measure for a pixel is computed by taking the average absolute value of pixel values in a square window of a convolution image. Laws provides sets of five element (L5, E5, S5, W5, and R5) and seven element (L7, E7, S7, W7, and R7) one-dimensional convolution kernels [22]. The names of the kernels are mnemonics for Level, Edge, Spot, Wave, and Ripple. Two-dimensional kernels are created by taking the outer product of any two of the one-dimensional kernels. After evaluating all 30 possible 5 by 5 and 7 by 7 two-dimensional kernels, we utilized a single Laws' texture energy measure derived from the L7 and E7 one-dimensional kernels. We use a 5 millimeter square window for the averaging step. This particular feature seems to respond well to region (i.e. mass) boundaries.

Roughness texture measure. This texture measurement computes the number of extrema per unit area [23], or the relative extrema density. Our definition of an extrema is slightly modified from that described in [23]. An extrema point is the location of a local minimum or maximum intensity value along a row or column. Thus, local extrema are computed separately for each row and column of the image. A run of pixels with the same intensity value at a local extrema is counted only as a single extrema point located at the midpoint of the run. To reduce the effects of noise, a local extrema is only counted if its relative difference from the previous extrema (in the row or column) exceeds a small threshold. The roughness texture measure at a pixel is computed for a 5 millimeter square window as the ratio of the total number of pixels in the window to the total number of relative extrema (computed in both directions) in the window. A rougher region will have more extrema and a higher feature value than a smooth region.

2.1.2 Classification

After examining many methods of statistical classification, including well-known Bayesian, nearest-neighbor, and neural network implementations, we settled on a decision tree classifier to perform the pixel-level classification. As we discuss shortly, we use well over 300,000 samples to train a classifier. With this much training data, the excessive training time required of typical neural network classifiers was prohibitive. Similarly, the test mode for nearest-neighbor classifiers is computationally too expensive since an unknown sample must be compared to every training sample in order to make a classification. The Bayesian classifiers utilize probability density functions which assume the data has a Gaussian distribution. Since we are attempting to detect several different types of masses, the wide range of visual characteristics they exhibit are not well modeled by a single Gaussian distribution.

Binary decision tree [24] (BDT) classification methods provide a means of approximating the optimal Bayes classification rule for a given situation. A BDT is simply an ordered list of binary threshold operations on the feature vectors, organized as a tree. At each node, one of the features in a vector is compared to a threshold, which moves the vector down the appropriate branch of the tree. This continues until it arrives at a terminal node which assigns a classification. The decision trees are grown automatically from the training data by recursive reduction of impurity. The control parameters at each node are chosen by simply determining the feature and threshold which best separate the current data. This process

is repeated, recursively partitioning the remaining training samples, until some stopping criteria is met. In our BDT implementation, quality of separation for a given feature and threshold is determined by a class separation measure, ORT, proposed in [25]. We use an iterative growing and pruning algorithm [26] for decision tree construction.

As noted above, we have defined the detection process as a 2-class problem. Pixels are classified as either normal or abnormal. That is, we made no attempt to assign different types of abnormalities to separate classes. This helps explain why the simple Gaussian assumption for a Bayesian classifier was not adequate. In decision trees, the leaf nodes may be seen as associating a probability with each class. The probability is computed from the training samples that fall into the leaf after the tree has been grown and pruned. For example, a leaf node may contain 80 training samples from class 1, and 20 training samples from class 2. During classification, we can say an unknown sample that falls into this leaf has an 80% probability of belonging to class 1, or a 20% probability of belonging to class 2. Thus, the probability of suspiciousness for a breast tissue pixel is computed in this manner.

2.2   Experimental Data

To evaluate performance of the software, a database of 117 digitized mammogram cases was utilized. Most are standard 4-view cases, 2 images of each breast, although a few cases had only 2 images as they were from mastectomy patients. In all, there are 465 images, 129 of which had at least 1 biopsy-proven abnormal mass. For some cases, the abnormality was not visible in one of the views, and so the database contains 137 visible lesions. There are three different types of lesions in the database: circumscribed masses, ill-defined masses, and spiculated masses. Ground truth in the form of an outline of the lesion boundary and an indication of benign or malignant pathology is provided for each mass. Ground truth was established by an expert mammographer. Approximately half of the lesions are malignant, and the other half are benign.

This set of images is publicly available as volume special_01 of the Digital Database for Screening Mammography [27, 28] at the University of South Florida (WWW address marathon.csee.usf.edu). The mammograms were digitized at a spatial resolution of 100 microns per pixel, and a grey level resolution of 12 bits per pixel. More information is available at the website.

A couple of standard preprocessing operations are applied to every image. First, the breast tissue was segmented from the background. This reduces processing time since feature extraction is only performed at pixels corresponding to breast tissue, and it also prevents background pixels from affecting feature values for breast tissue pixels. Second, the images were scaled to a spatial resolution of 300 microns per pixel. This also reduces processing time yet still permits reliable detection of the smaller masses in the database.

2.3   Train and Test Protocol

In order to make efficient use of all available mammogram data, a v-fold cross-validation method was employed for training and testing. Basically, the 117 cases were divided into 6 sets of data. Five of the sets contained 20 cases each, and the last test set contained 17 cases (for a total of 117 cases). The data sets are mutually exclusive with regards to case

selection. So, each case belongs to one and only one data set. Results for any one data set are obtained by using the remaining five data sets for training purposes. Thus, at no time is any data used for training and testing in the same set of experiments. The advantage of the cross-validation method is that all available data can be used independently for training *and* testing without biasing the results.

To train the decision tree classifier, pixel locations are randomly sampled from within the normal and abnormal regions of each training image. Since we have more normal images than abnormal images, 500 normal samples are obtained from every image, and 1700 abnormal samples are obtained from those images containing an abnormality. If an abnormality contains less then 1700 pixel, then all pixels belonging to the abnormality are sampled. Considering that each time we train the classifier we use samples taken from about 97 cases (117 total cases minus 20 cases held out for testing), or around 388 images (4 images per case), there is an abundance of training data at this stage of processing. In general, we have almost 200,000 samples from normal tissue and around 130,000 samples from abnormal tissue each time the decision tree classifier is trained.

For a given set of training images, the seven features described above are extracted from the randomly sampled pixel locations, and a decision tree is grown. A parameter in the decision tree algorithm ensures that leaf nodes in the tree will contain at least 50 samples. With the abundance of training data, this prevents the decision tree algorithm from over-fitting, and leads to better generalization. For a given test image, feature measurements are extracted, and each pixel is assigned a probability of suspiciousness in the manner described above. The resulting probability image is smoothed with a 5 mm uniform kernel to obtain a consensus among neighboring pixels, and thresholded to eliminate pixels with a low probability of suspiciousness. Next, pixels are grouped into 4-connected regions, and very small regions (less than 10 pixels) are eliminated. Finally, cross-hairs are overlaid on the original mammogram image at the centroid of any remaining regions.

2.4   Performance Evaluation

A couple of basic assumptions concerning the intended use of mammogram image analysis software dictate what constitutes acceptable performance. First, since the system is meant to prompt radiologists by directing their attention to suspicious regions on a mammogram, accurate segmentation of potential lesions is not required. We only need to place a prompt, such as the cross-hairs, somewhere within the boundary of a lesion. Therefore, any prompt lying within the boundary of a lesion is considered a true positive detection, while all other prompts are considered false positives. Second, most lesions are visible in two views since there are two views of each breast in a screening case. Therefore, it is acceptable to detect a mass only in one of the two views.

The sensitivity of the detection routine can be adjusted by varying the threshold applied to the probability image (see Figure 6). A lower threshold, results in a more sensitive the algorithm, and the detection rate increases. However, this higher sensitivity comes at a cost of more false positive prompts per image. To examine the effect of this threshold, probability images were thresholded over the range 0.5 to 1.0 in increments of 0.005, and the results were collected.

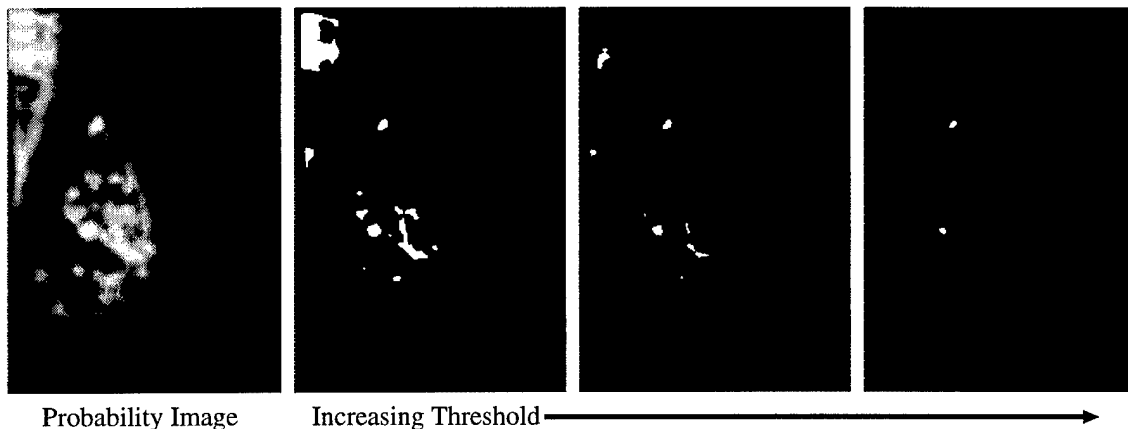Probability Image   Increasing Threshold ➞

Figure 6: Example of various threshold levels applied to the probability image shown in Figure 1.

2.5 Results

For an average of about 2 false positive detections per image, our algorithm detected 70% (or 96) of the 137 lesions. In a CAD second reader scenario it is only necessary to detect a lesion in at least one of the two views. Using this criteria, our algorithm detected 90% of the masses in at least one view, and perhaps more importantly 97% of the malignant lesions were found in at least one view. Approximately 20% of the images had no detections, meaning the radiologist would not have to re-examine those images. Another set of results obtained by varying the threshold on the probability images shows that a lesion sensitivity (in at least on view) of 70% can be achieved with an average of one false positive prompt per image.

3 Conclusions and Recommendations

In conclusion, we have developed image analysis software which is capable of detecting mammographic abnormalities which would be used in a second reader scenario to prompt a radiologist to more carefully analyze suspicious regions in the mammogram. The automated prompting and the additional information provided by computerized image analysis should result in greater repeatability and uniformity in the standard of care. As several recent studies have indicated, it should also result in some increase in sensitivity for a given level of specificity; that is, fewer missed cancers with the same biopsy rate. The cancers detected as a result of the increased sensitivity would then be treated earlier and thus less expensively and with a higher cure rate.

Based on our experiences associated with this project, we can make several recommendations for other researchers in this field.

From a high-level view, our approach to detecting abnormalities in mammograms consists of two steps: pixel-level feature extraction, and statistical classification. A good portion of our early research efforts concentrated on developing better methods of statistical classification [29, 30, 31]. While these efforts did lead to improved classification accuracy at the

10

pixel level, the computational complexity of the algorithms make them not especially well-suited for a problem of this magnitude (hundreds of thousands of training samples). Later work stressed the importance of engineering problem-specific feature extraction algorithms [15]. It was here that the most significant improvements were realized. Therefore, we would recommend other researchers concentrate on designing features that are highly sensitive and specific.

Other areas of research which may lead to significant improvements in performance are preprocessing routines. One example is noise equalization [32] which can lead to more reliable and robust feature extraction routines. A visual examination of our feature images shows there can be undesirable responses near the breast roll-off region in the mammograms. This roll-off region is near the breast boundary where the thickness of the breast gradually decreases because of the way the breast is compressed during imaging. A preprocessing method to correct the breast thickness roll-off [33] should lead to improved feature extraction, and ultimately better overall performance. A number of false positive detections can be attributed to an undesirable response from the feature extraction routines caused by the pectoral muscle in the MLO views. While it may not be desirable to segment this region out of the image since masses behind the pectoral muscle can be visible, some form of thickness equalization may lead to improved performance.

In this work, the region-level analysis is quite simple. We simply remove very small regions. This is, in part, due to the results produced by our detection algorithm. The goal was not an accurate segmentation, but rather accurate localization of suspicious regions. As a result, the regions segmented from the image generally correspond to the central region of the mass, and not much of the lesion boundary is captured. More sophisticated post-processing and region-level analysis should result in a reduction of false positives. One possibility is to perform region growing on the segmented blobs, followed by region classification based on size, shape, and other properties more appropriate for describing regions (i.e. connected groups of pixels).

4   Other Work Related to This Project

The following publications are the result of this project.

1. Woods, K.S., Kegelmeyer Jr, W.P., and Bowyer, K.W. Combination of multiple classifiers using local accuracy estimates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19** (4), 405-410, (April 1997).

2. Woods, K.S., and Bowyer, K.W. Generating ROC curves for artificial neural networks, *IEEE Transactions on Medical Imaging*, **16** (3), June 1997.

3. Woods, K.S., Kegelmeyer Jr, W.P., and Bowyer, K.W. Combination of multiple classifiers using local accuracy estimates, Proceedings of the *1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, San Francisco, California, (June 1996).

4. Woods, K.S., and Bowyer, K.W. A General View of Detection Algorithms, in Digital Mammography '96, (*Proceedings of the Third International Conference on Digital Mammog-*

raphy held in Chicago, Illinois, June 1996), K. Doi, M.L. Giger, R.M. Nishikawa, and R.A. Schmidt, editors, Elsevier Science, Amsterdam, 1996, 385-390.

5. Woods, K.S., and Bowyer, K.W. Computer detection of stellate lesions, in <u>Digital Mammography</u>, (*Proceedings of the Second International Workshop on Digital Mammography*, held in York, United Kingdom, July 1994), A.G. Gale, S.M. Astley, D.R. Dance, and A.Y. Cairns, editors, Elsevier Science, Amsterdam, 1994, 221-229.

6. Bowyer, K., Kopans, D., Kegelmeyer Jr, W.P., Moore, R., Sallam, M., Chang, K., and Woods, K. The digital database for screening mammography, in <u>Digital Mammography</u>, (*Proceedings of the Second International Workshop on Digital Mammography*, held in York, United Kingdom, July 1994), A.G. Gale, S.M. Astley, D.R. Dance, and A.Y. Cairns, editors, Elsevier Science, Amsterdam, 1994, 431-434.

7. Woods, K.S., and Bowyer, K.W. Generating ROC curves for artificial neural networks, *Seventh Annual IEEE Symposium on Computer-Based Medical Systems*, Winston-Salem, North Carolina, (June 1994), 201-206.

8. Solka, J.L., Poston, W.L., Priebe, C.E., Rogers, G.W., Lorey, R.A., Marchette, D.J., Woods, K.S., and Bowyer, K.W. The detection of microcalcifications in mammographic images using high dimensional features, *Seventh Annual IEEE Symposium on Computer-Based Medical Systems*, Winston-Salem, North Carolina, (June 1994), 139-145.

9. Sallam, M., Bowyer, K., Woods, K., Kopans, D., Moore, R., and Kegelmeyer Jr, W.P. The digital database for screening mammography (DDSM): lessons learned, Radiological Society of North America (RSNA) 1997 (Abstract).

References

[1] American Cancer Society, "Cancer facts and figures," Atlanta, GA, 1993.

[2] D. B. Kopans, *Breast Imaging*. Philadelphia, PA: J.B. Lippincott Company, 1989.

[3] C. Beam, P. Layde, and D. Sullivan, ""Variability in the interpretation of screening mammograms by US radiologists"," *Archives of Internal Medicine*, vol. 156, pp. 209–213, Jan. 1996.

[4] D. B. Kopans, "Discriminating analysis uncovers breast lesions," *Diagnostic Imaging*, pp. 94–101, September 1991.

[5] J. E. Martin, M. Moskowitz, and J. R. Milbrath, "Breast cancer missed by mammography," *AJR*, vol. 132, pp. 737–739, 1979.

[6] C. J. Baines, D. V. MacFarland, and A. B. Miller, "The role of the reference radiologist: estimates of interobserver agreeement and potential delay in cancer detection in the national breast screening study," *Investigative Radiology*, vol. 25, pp. 971–976, 1990.

[7] R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology*, vol. 184, pp. 613–617, 1992.

[8] J. A. Harvey, L. L. Fajardo, and C. A. Inis, "Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation," *AJR*, vol. 161, pp. 1167–1172, 1993.

[9] J. A. vanDijck, A. L. Verbeek, J. H. Hendriks, and R. Holland, "The current detectability of breast cancer in a mammographic screening program," *Cancer*, vol. 72, pp. 1933–1937, 1993.

[10] W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. H. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, pp. 331–337, 1994.

[11] S. Astley, I. Hutt, S. Adamson, P. Miller, P. Rose, C. Boggis, C. J. Taylor, T. Valentine, J. Davies, and J. Armstrong, "Automation in mammography: computer vision and human perception," in *Proceedings of the SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, vol. 1905, (San Jose, CA), pp. 716–730, Jan 31 - Feb 4 1993.

[12] H. P. Chan, K. Doi, and C. J. Vyborny, "Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis," *Investigative Radiology*, vol. 25, pp. 1102–1110, 1990.

[13] R. A. Schmidt, R. M. Nishikawa, K. Schreibman, M. L. Giger, K. Doi, J. Papaioannou, P. Lu, J. Stucka, and G. Birkhahn, "Computer detection of lesions missed by mammography," in *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, vol. 1069 of *International Congress Series*, (York, England), pp. 289–294, Elsevier Science B. V., July 10-12 1994.

[14] I. W. Hutt, S. M. Astley, and C. R. M. Boggis, "Prompting as an aid to diagnosis in mammography," in *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, vol. 1069 of *International Congress Series*, (York, England), pp. 389–398, Elsevier Science B. V., July 10-12 1994.

[15] K. S. Woods and K. W. Bowyer, "A general view of detection algorithms," in *Proceedings of the Third International Conference on Digital Mammography*, (Chicago, Illinois), pp. 385–390, Elsevier Science B. V., June 1996.

[16] S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Transactions on Medical Imaging,* vol. 8, pp. 377–386, December 1989.

[17] W. P. Kegelmeyer, Jr., "Computer detection of stellate lesions in mammograms," in *Proceedings of SPIE Conference on Biomedical Image Processing*, 1992.

[18] W. P. Kegelmeyer, Jr., "Evaluation of stellate lesion detection in a standard mammogram data set," in *Proceedings of the SPIE/IS&T Symposium on Electronic Imaging Science and Technology*, vol. 1905, (San Jose, CA), pp. 787–798, Jan 31 - Feb 4 1993.

[19] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 891–906, September 1991.

[20] J. L. Solka, C. E. Priebe, and G. W. Rogers, "An initial assessment of discriminant surface complexity for power law features," *Simulation*, vol. 58, no. 5, pp. 311–318, 1992.

[21] S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 518–523, July 1984.

[22] K. I. Laws, *Textured Image Segmentation*. PhD thesis, University of Southern California School of Engineering, 1980.

[23] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision: Volume 1*. Addison-Wesley, 1992.

[24] L. Breiman, J. H. Friedman, R. A. Olsen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group, 1984.

[25] U. M. Fayyad and K. B. Irani, "The attribute selection problem in decision tree generation," in *Proceedings of AAAI*, (San Jose, CA), pp. 104–110, July 1992.

[26] S. B. Gelfand, C. S. Ravishankar, and E. J. Delp, "An iterative growing and pruning algorithm for classification tree design," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 163–174, February 1991.

[27] K. Bowyer, D. Kopans, W. P. K. Jr, R. Moore, M. Sallam, K. Chang, and K. Woods, "The digital database for screening mammography," in *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, vol. 1069 of *International Congress Series*, (York, England), pp. 431–434, Elsevier Science B. V., July 10-12 1994.

[28] M. Sallam, K. Bowyer, K. Woods, D. Kopans, R. Moore, and W. P. K. Jr, "The digital database for screening mammography (ddsm): lessons learned," in *1997 Scientific Program*, vol. 205 (P), (Chicago, Illinois), Radiological Society of North America (RSNA), July 10-12 1997.

[29] K. S. Woods, W. P. K. Jr, and K. W. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 405–410, April 1997.

[30] K. S. Woods and K. W. Bowyer, "Generating roc curves for artificial neural networks," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 377–386, June 1997.

[31] K. S. Woods, W. P. K. Jr, and K. W. Bowyer, "Combination of multiple classifiers using local accuracy estimates," in *Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, (San Fransisco, California), June 1996.

[32] N. Karssemeijer, *Adaptive Noise Equalization and Recognition of Microcalcification Clusters in Mammograms*, vol. 9 of *Machine Perception Artificial Intelligence*, pp. 148–166. Singapore: World Scientific Publishing Company, 1994.

[33] J. W. Byng, J. P. Critten, and M. J. Yaffe, "Thickness-equalization processing for mammographic images," *Radiology*, vol. 203, pp. 564–568, 1997.